

1. Bagaimanakah kualitas butir soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman meliputi aspek materi?
2. Bagaimanakah kualitas butir soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman meliputi aspek bahasa?
3. Bagaimanakah kualitas butir soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman meliputi aspek konstruksi?
4. Bagaimanakah validitas isi soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman?
5. Bagaimanakah estimasi reliabilitas soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman?
6. Bagaimanakah tingkat kesulitan butir soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman?
7. Bagaimanakah daya beda soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten Sleman?
8. Bagaimanakah *pseudo guessing* soal penilaian akhir semester (PAS) mata pelajaran IPS Kelas VII SMP Negeri di Kabupaten?

BAB III METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini merupakan penelitian analisis dokumen yang menggunakan pendekatan kuantitatif deskriptif untuk mengetahui kualitas perangkat tes Penilaian Akhir Semester (PAS) mata pelajaran IPS SMP kelas VII tahun ajaran 2016/2017 di tingkat Kabupaten Sleman. Tujuan yang ingin dicapai adalah untuk memperoleh butir-butir soal mata pelajaran IPS tingkat SMP yang memenuhi kriteria validitas dan reliabilitas, sehingga butir soal tersebut dapat digunakan sebagai bank soal mata pelajaran IPS SMP. Sumber data dalam penelitian ini berupa respon siswa pada PAS IPS SMP kelas VII tahun ajaran 2016/2017 di tingkat Kabupaten Sleman yang selanjutnya dianalisis.

Penelitian ini menggunakan telaah secara teoritis oleh *expert judgement* dan penelaahan butir soal dilanjutkan dari hasil respon jawaban siswa SMP kelas VII secara empiris, dengan bantuan program komputer *Bilog MG* untuk teori respon butir. Hasil analisis diharapkan dapat diperoleh butir-butir soal yang berkualitas baik dan memenuhi kriteria.

B. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Kabupaten Sleman. Alokasi waktu untuk penelitian ini adalah mulai dari bulan November 2017 sampai Desember 2018.

C. Subjek dan Responden

Subjek dalam penelitian ini adalah siswa SMP negeri kelas VII di Kabupaten Sleman. Jumlah SMP Negeri di Kabupaten Sleman sebanyak 54 sekolah, dengan

jumlah seluruh siswa kelas VII dari 54 sekolah tersebut sebanyak 7708 siswa. Berdasarkan jumlah subjek tersebut, maka jumlah minimal responden yang digunakan dalam penelitian ini menurut tabel Krejcie dan Morgan dengan taraf signifikansi 5% sebanyak adalah 367 siswa.

Pengambilan responden dilakukan dengan teknik *stratified random sampling*. Adapun strata dalam penelitian ini yaitu SMP N di Kabupaten Sleman dikelompokkan berdasarkan nilai PAS tahun 2016/2017 menjadi tiga tingkatan yaitu tinggi, sedang, dan rendah. Penggunaan teknik ini dilakukan agar responden yang digunakan dapat merepresentasikan keadaan subjek, sehingga hasil penelitian ini dapat digeneralisasi di Kabupaten Sleman. Responden penelitian dapat dilihat pada Tabel 4 berikut.

Tabel 4. Responden Penelitian

Nama Sekolah	Jumlah Siswa	Strata
SMP Negeri 4 Pakem	157	Tinggi
SMP Negeri 1 Godean	127	Tinggi
SMP Negeri 4 Depok	127	Tinggi
SMP Negeri 1 Ngaglik	176	Sedang
SMP Negeri 2 Kalasan	187	Sedang
SMP Negeri 3 Mlati	122	Sedang
SMP Negeri 5 Sleman	126	Rendah
SMP Negeri 2 Gamping	192	Rendah
SMP Negeri 2 Prambanan	121	Rendah
Jumlah	1335	

Sumber (Peneliti, 2018)

Tabel 4 di atas menunjukkan bahwa terdapat 9 SMP Negeri di Kabupaten Sleman yang dijadikan sebagai responden penelitian ini. Kesembilan sekolah tersebut terbagi dalam tiga strata yaitu tiga sekolah termasuk kategori “tinggi”, tiga

sekolah kategori “sedang”, dan pada kategori “rendah” terdapat tiga sekolah. Jumlah responden yang digunakan sebanyak 1335 siswa.

D. Teknik dan Instrumen Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini adalah dokumentasi. Data dalam penelitian ini dikumpulkan melalui lembar telaah dan lembar validasi oleh *expert judgement*, serta hasil analisis empiris dari respons jawaban siswa pada PAS IPS SMP kelas VII. Adapun instrumen pengumpulan data pada penelitian ini adalah sebagai berikut.

1. Kisi-kisi Soal PAS IPS Kelas VII SMP Di Kabupaten Sleman Tahun Ajaran 2016/2017 (dalam Lampiran 1).
2. Perangkat tes PAS mata pelajaran IPS SMP kelas VII berupa soal pilihan ganda (dalam Lampiran 2).
3. Lembar telaah dan lembar validasi yang telah dianalisis oleh *Ekspert Judgement* (dalam Lampiran 9, 10, 15).
4. *Softcopy* lembar jawaban siswa dalam bentuk program excel yang didapat dari Media Scanner terhadap perangkat tes PAS mata pelajaran IPS SMP tahun ajaran 2016/2017 (dalam Lampiran 16).

E. Teknik Analisis Data

Analisis dalam penelitian ini menggunakan analisis kualitatif dan kuantitatif. Analisis kualitatif bertujuan untuk menentukan kualitas soal secara teoritik sedangkan analisis kuantitatif bertujuan untuk menentukan kualitas soal secara empiris. Data yang dianalisis secara kualitatif meliputi data yang diperoleh dari

lembar validasi, lembar telaah kualitatif, dan data yang dianalisis secara kuantitatif diperoleh dari lembar soal UAS IPS kelas VII.

1. Analisis Data Kualitatif

Telaah butir soal yang disusun merupakan penilaian (*judgement*) pada setiap butir soal dengan kriteria telaah dari aspek materi, konstruksi dan bahasa. Telaah dilakukan oleh tiga orang ahli untuk aspek materi & konstruksi, yaitu dua orang dosen IPS, dan satu guru IPS kelas VII yang telah memiliki pengalaman dan kompeten. Setiap ahli mengerjakan perangkat tes secara terpisah, dengan cara mengisi lembar telaah yang dibuat sesuai panduan yang telah ditetapkan. Keputusan penelitian didasarkan pada data dukungan dari ketiga penelaah, dengan ketentuan keputusan sebagai berikut.

- a. Baik, apabila keseluruhan kriteria penelaahan butir pada aspek materi, konstruksi dan bahasa semuanya sesuai kaidah yang ditentukan, dan didukung oleh semua penelaah (81-100%).
- b. Cukup baik, apabila keseluruhan kriteria penelaahan butir pada aspek materi sesuai dengan kaidah yang ditentukan, sebanyak-banyaknya ada satu kriteria pada aspek konstruksi dan satu kriteria pada aspek bahasa yang tidak sesuai dengan kaidah yang ditentukan, minimal didukung oleh dua orang penelaah (70-80%).
- c. Tidak baik, apabila kunci jawaban salah, atau terdapat kriteria pada aspek materi yang tidak sesuai dengan yang ditentukan, lebih dari satu kriteria pada aspek konstruksi dan lebih dari satu kriteria pada aspek bahasa yang tidak

sesuai dengan kaidah yang telah ditetapkan, minimal didukung oleh dua orang penelaah (<69%).

2. Analisis Data Kuantitatif

Analisis secara kuantitatif dilakukan dengan menentukan validitas isi, reliabilitas dan analisis butir soal secara empiris dengan pendekatan teori respon butir (IRT)

a. Validitas Isi

Cara menentukan validitas isi, yaitu validitas isi disahkan oleh 3 orang ahli dan validitas isi dilakukan secara kuantitatif dengan melihat masukan sebagai perbaikan butir instrumen. Ketiga ahli tersebut melakukan penilaian terhadap instrumen tes. Untuk membuktikan validitas isi melalui kesepakatan ini, dapat digunakan indeks validitas yang diusulkan oleh Aiken (Retnawati, 2016: 18). Penentuan validitas isi merupakan penilaian dan masukan dari ketiga ahli sebagai validator instrumen yang diuji cobakan. Perhitungan validasi isi menggunakan rumus *Aiken's* sebagai berikut:

$$V = \frac{\sum s}{n(c - 1)}$$

dimana $s = r - I_o$

Keterangan:

- V : Indeks kesepakatan rater mengenai validitas butir
- s : Skor yang ditetapkan setiap rater dikurangi skor terendah dalam kategori yang dipakai
- r : Skor kategori pilihan rater
- I_o : Skor terendah dalam kategori penyekoran
- n : Banyaknya rater
- c : Banyaknya kategori yang dapat dipilih rater

Setelah menghitung koefisien, jika indeks kesepakatan $\leq 0,4$ maka dikatakan validitasnya rendah, sedangkan indeks kesepakatannya diantara 0,4-0,8 dikatakan validitasnya sedang, dan jika indeks kesepakatan $> 0,8$ maka dikatakan sangat valid (Retnawati, 2016: 19).

b. Estimasi Reliabilitas

Teknik analisis data untuk estimasi reliabilitas menggunakan teknik estimasi konsistensi internal dengan formula diperoleh menggunakan Formula *Cronbach Alpha*, yaitu sebagai berikut:

$$\alpha = \left[\frac{k}{(k-1)} \right] \left[1 - \frac{\sum \alpha i^2}{\alpha t^2} \right]$$

Keterangan

α : Koefisien Reliabilitas instrumen

k : Jumlah Item Dalam Tes

$\sum \alpha i^2$: Jumlah Varians butir Instrumen

αt^2 : Varians skor total

c. Analisis Butir Soal

Analisis butir soal pada penelitian ini menggunakan teori respon butir. Data yang diperoleh terlebih dahulu dilakukan uji asumsi sebelum dianalisis dengan pendekatan teori respon butir. Terdapat dua asumsi yang dimiliki teori respon butir yaitu unidimensi dan independensi lokal (Hambleton et al, 1991: 9-12; Mardapi, 2012: 201). Asumsi unidimensi merupakan hal pertama yang harus dipenuhi dalam analisis teori respon butir. Asumsi unidimensi dilakukan dengan melakukan analisis faktor eksploratori. Asumsi kedua yaitu independensi lokal, jika faktor-faktor yang mempengaruhi prestasi konstan maka respon subyek

terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain (Retnawati, 2014: 3).

Pada teori respon butir terdapat tiga model logistik satu parameter, model logistik dua parameter dan model logistik tiga parameter. Model logistik satu parameter (1 PL) melibatkan satu parameter butir yaitu tingkat kesukaran (b). Model logistik dua parameter (2 PL) melibatkan dua parameter butir yaitu tingkat kesukaran (b) dan daya beda (a). Sedangkan model logistik tiga parameter (3 PL) melibatkan dua parameter butir yaitu tingkat kesukaran (b), daya beda (a) dan tebakan semu/ *pseudue guessing* (c). Analisis pada penelitian ini dilakukan dengan menggunakan bantuan program *Bilog MG*. Model logistik yang digunakan pada analisis butir soal dalam teori respon butir adalah sebagai berikut:

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}; \quad 1PL$$

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}; \quad 2PL$$

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{1,7a_i(\theta-b_i)}}{1 + e^{1,7a_i(\theta-b_i)}}; \quad 3PL$$

Keterangan:

$P_i(\theta)$: probabilitas menjawab benar butir i oleh peserta berkemampuan

θ : kemampuan peserta

e : bilangan transenden yang besarnya mendekati 2,718

n : banyaknya butir dalam tes

a_i : daya pembeda butir i

b_i : tingkat kesukaran butir i

c_i : tebakan (*pseudo guessing*) butir i

D : faktor skala (1,7)

(Hambleton et al., 1991)

Beberapa aspek atau karakteristik butir yang perlu diperhatikan dalam teori respon butir, yaitu:

1) Daya Beda (a)

Daya beda (a) merupakan kemampuan butir yang dengan skornya dapat membedakan peserta dari kelompok tinggi dan kelompok rendah. Indeks tersebut terkadang disebut *slope*, karena menunjukkan seberapa tajam kemungkinan perubahan respon yang menjawab benar seperti kemampuan atau kenaikan trait (ciri). Menurut Hambleton et al. (1985: 36) indeks daya beda berada di antara 0 – 2.

2) Tingkat Kesukaran (b)

Indeks kesukaran butir (b) diukur melalui skor butir yang dihasilkan oleh jawaban sejumlah peserta tes. Semakin banyak peserta tes mampu menjawab soal tes yang diberikan maka tingkat kesukaran tes tersebut semakin rendah dan sebaliknya. Butir soal yang baik terletak pada interval $-2 \leq \theta \leq 2$ (Hambleton et al., 1991: 13). Nilai b mendekati -2 maka mengidentifikasikan bahwa butir tersebut semakin mudah, dan jika nilai b mendekati +2 maka mengidentifikasikan bahwa butir soal semakin sulit.

3) Pseudo Guessing (c)

Parameter c adalah *lower asymptot* dari kurva karakteristik butir. Hambleton et al. (1985: 38) menyebutkan bahwa parameter c menunjukkan peluang peserta tes dengan kemampuan rendah mampu menjawab butir dengan benar. Pada tes pilihan ganda yang terdiri dari k pilihan alternatif jawaban, maka besarnya parameter c terletak sekitar $\frac{1}{k}$.

4) Kecocokan Butir/ *Goodness of fit statistic*

Kecocokan suatu butir dengan model dapat dilihat dari nilai *chi*-kuadrat butir dibandingkan dengan nilai kritik distribusi *chi*-kuadrat sesuai dengan *dk* butir yang bersangkutan pada taraf signifikansi = 0,01 atau = 0,05. Butir dikatakan cocok model jika nilai *chi* kuadrat butir lebih kecil dari nilai distribusi *chi*-kuadrat pada nilai kritisnya atau jika menggunakan nilai signifikansi (*p*), butir dikatakan cocok model jika nilai signifikansi (*p*) lebih dari taraf signifikansi α .

5) Fungsi Informasi

Setiap butir soal memiliki fungsi informasi dan jumlahnya merupakan fungsi informasi tes tersebut (Hambleton et al., 1991: 94). Fungsi informasi paket tes akan tinggi jika butir penyusunnya mempunyai fungsi informasi yang tinggi pula. Jika fungsi informasi tes adalah $I(\theta)$ fungsi informasi butir $I_i(\theta)$, secara matematis fungsi informasi tes dapat ditulis (Hambleton et al., 1991: 94):

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

6) Kesalahan Pengukuran atau *Standard Error Measurement* (SEM)

Kesalahan *error* pengukuran menurut teori respons butir dinyatakan dengan *Standar Error of Measurement/ (SEM)*. Fungsi informasi dan SEM mempunyai hubungan yang berbanding terbalik kuadratik, yaitu semakin besar nilai fungsi informasi berarti SEM semakin kecil dan sebaliknya. Jika, fungsi informasi dinyatakan dengan $I(\theta)$ dan kesalahan baku pengukuran

dinyatakan dengan $SEM(\theta)$, bentuk hubungan keduanya dirumuskan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991: 94):

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$$